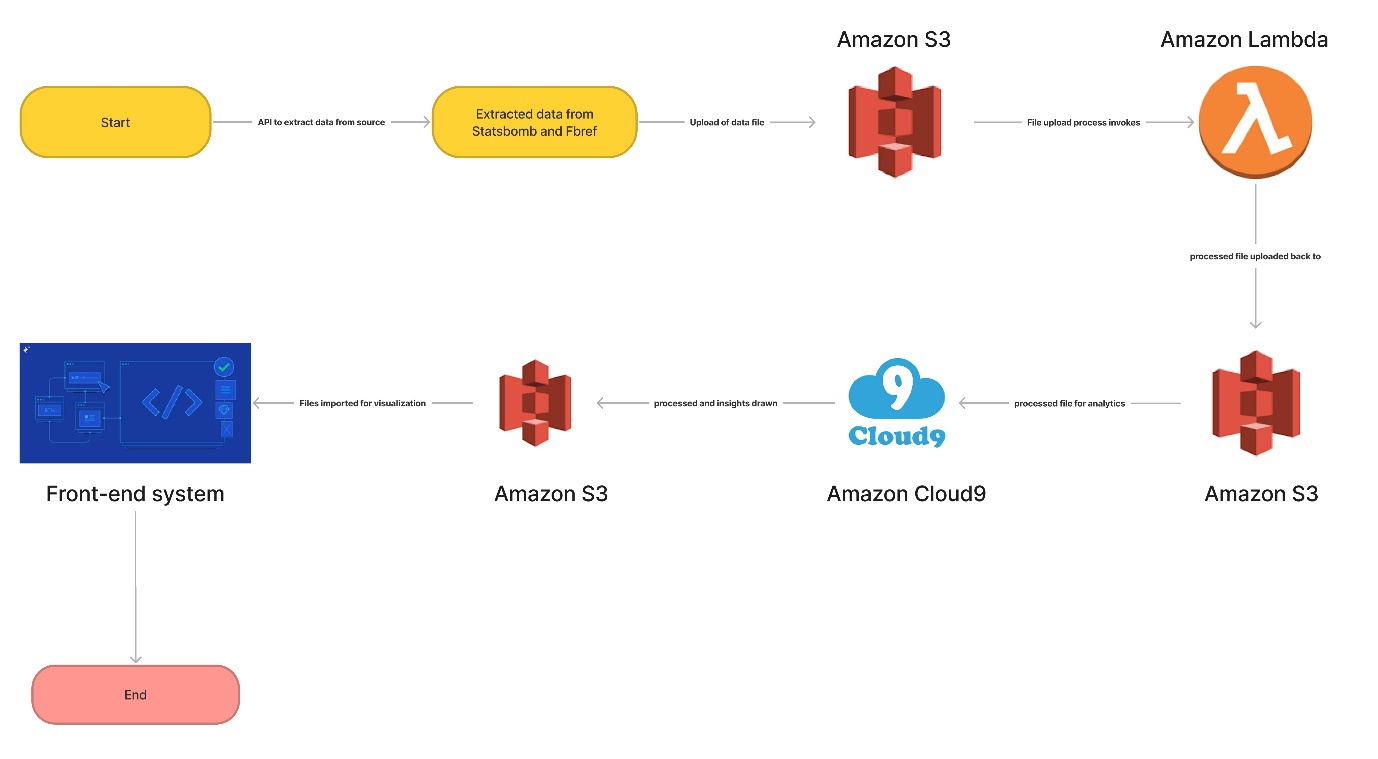
Data Pipeline Architecture



Extracted Data -> Amazon S3

Statsbomb has provided some great open data which can be extracted using the API provided as it is the simpler, quicker and easier way to access the open data. Statsbomb has provided an API to extract the open data with the use of library “statsbombpy”. Using this library, from Statsbomb, we can extract data of events in a specific match from specific season of a specific competition.

Fbref provides open-source data set of players report, we can just import the information of a specific player from their website into a csv file.

We can then store these extracted data into Amazon S3 environment

You can simply exchange data stored in Amazon Simple Storage Service (Amazon S3) with numerous apps. Each application, however, has its own set of needs and may demand a different representation of the data.

Amazon S3 -> Lambda function -> Amazon S3

AWS Lambda is an Amazon Web Services serverless computing service (AWS). AWS Lambda users write functions, which are self-contained programs written in one of the supported languages and runtimes, and upload them to AWS Lambda, which then executes them in a fast and flexible way.

Lambda functions may be used to do everything from providing web pages to processing data streams to using APIs and connecting with other AWS services.

S3 Object Lambda is a feature that lets you write your own code to treat data from S3 before sending it to an application. S3 Object Lambda integrates with current apps and utilizes AWS Lambda functions to handle and alter data as it is pulled from S3.

Once a file is uploaded to the S3 bucket, the lambda function will be synchronously run inline with a typical S3 GET request and begin processing the file's content; once done, the lambda function will upload the processed file back to the S3 bucket.

Amazon S3 -> AWS Cloud9 environment -> Amazon S3

AWS Cloud9 is a browser-based integrated development environment (IDE) that allows you to write, execute, and debug code. A code editor, debugger, and terminal are all included. Cloud9 is pre-installed with key tools for major programming languages such as JavaScript, Python, PHP, and more, so you don't have to install files or setup your development workstation to get started.

Cloud9 also offers a smooth development experience for serverless apps, allowing you to simply specify resources, debug, and switch between local and remote execution. You can share your development environment with your team fast using Cloud9, allowing you to pair program and track each other's inputs in real time.

Objects in an Amazon S3 bucket will be downloaded to a folder in your AWS Cloud9 environment from the AWS Cloud. After that, you may use analytics in Python to program and draw conclusions from this data file. Once this file has been processed, it may be returned to the Amazon S3 bucket.

The upload to Amazon S3 from AWS Cloud9 will be done using the Toolkit interface or a command to upload a file to a bucket.

Both methods will upload a file from the AWS Cloud9 environment to the AWS Cloud and save it as an S3 object. A file can be uploaded to a bucket or a folder that organizes the contents of that bucket.

Amazon S3 -> Front-end system

This pre-processed and analyzed final data set file will be picked up and utilized in the integration to the front-end system for visualization purposes.